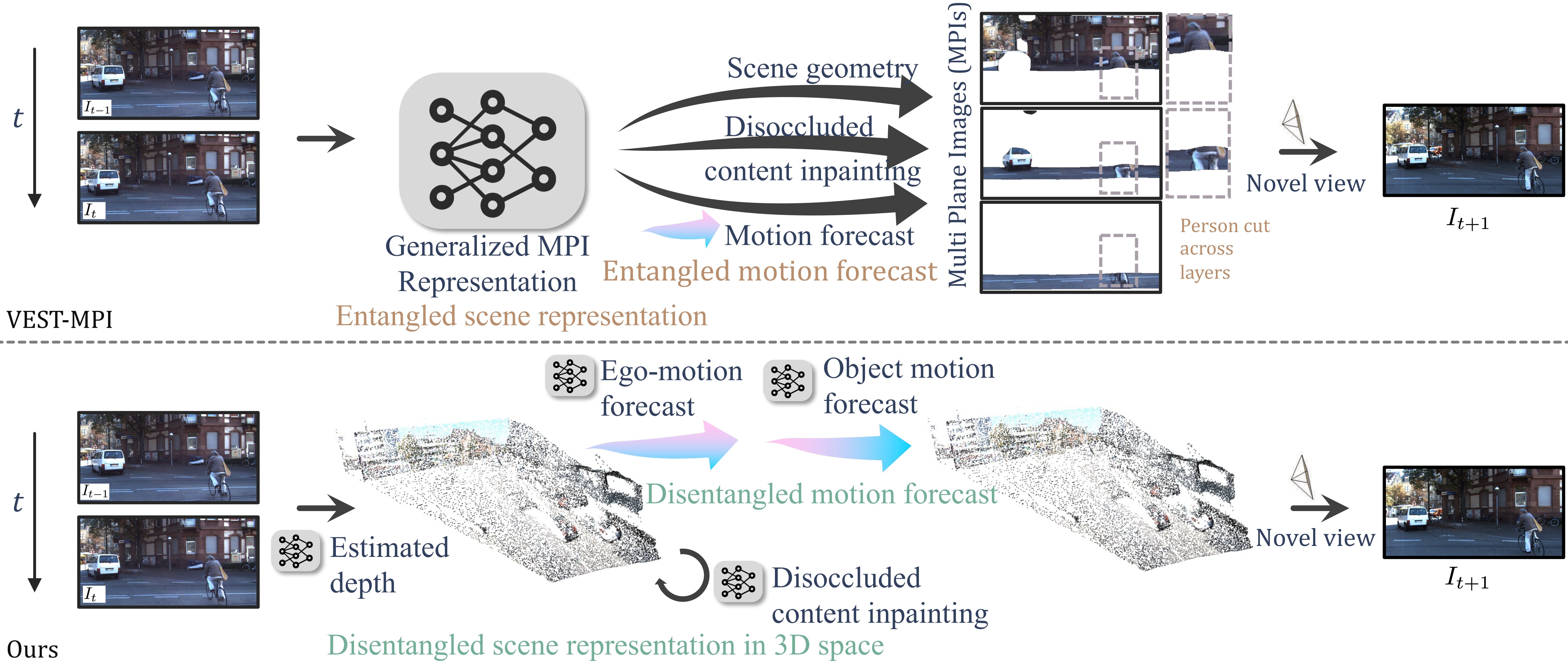
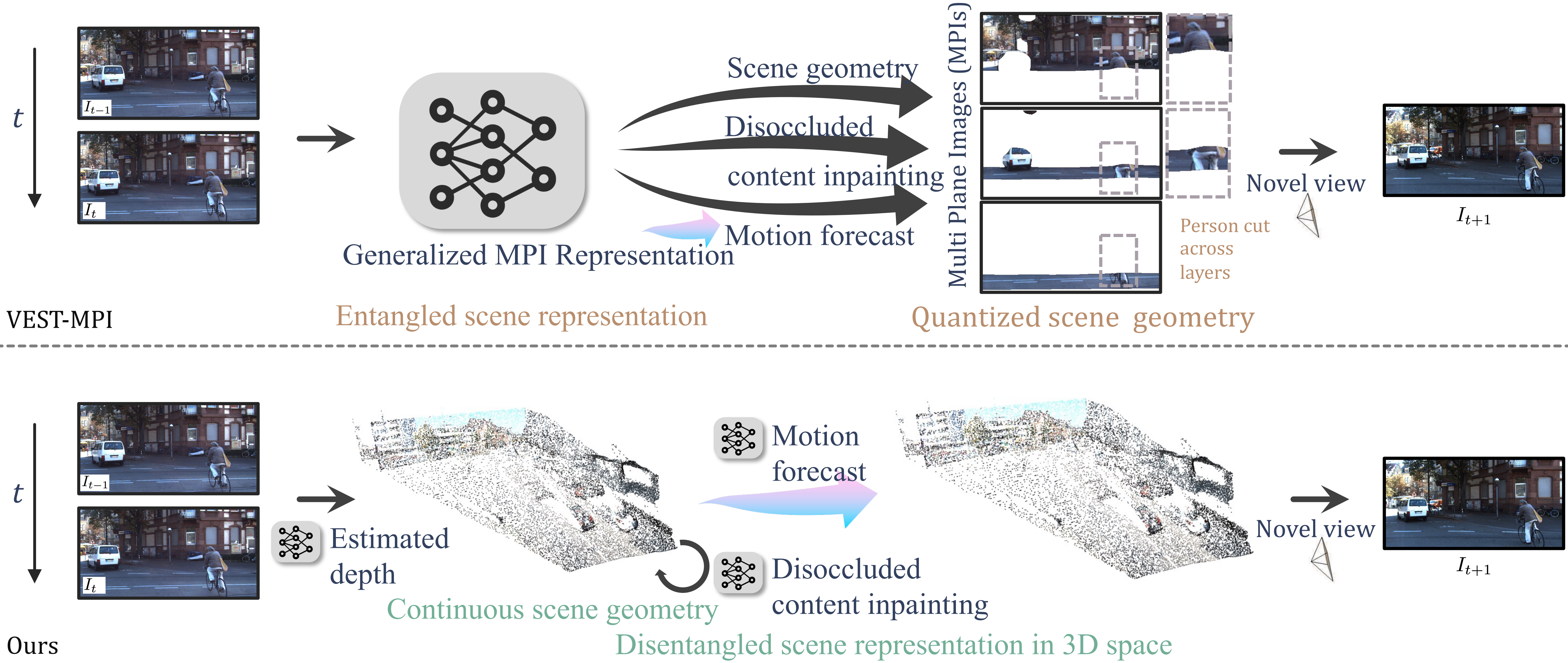
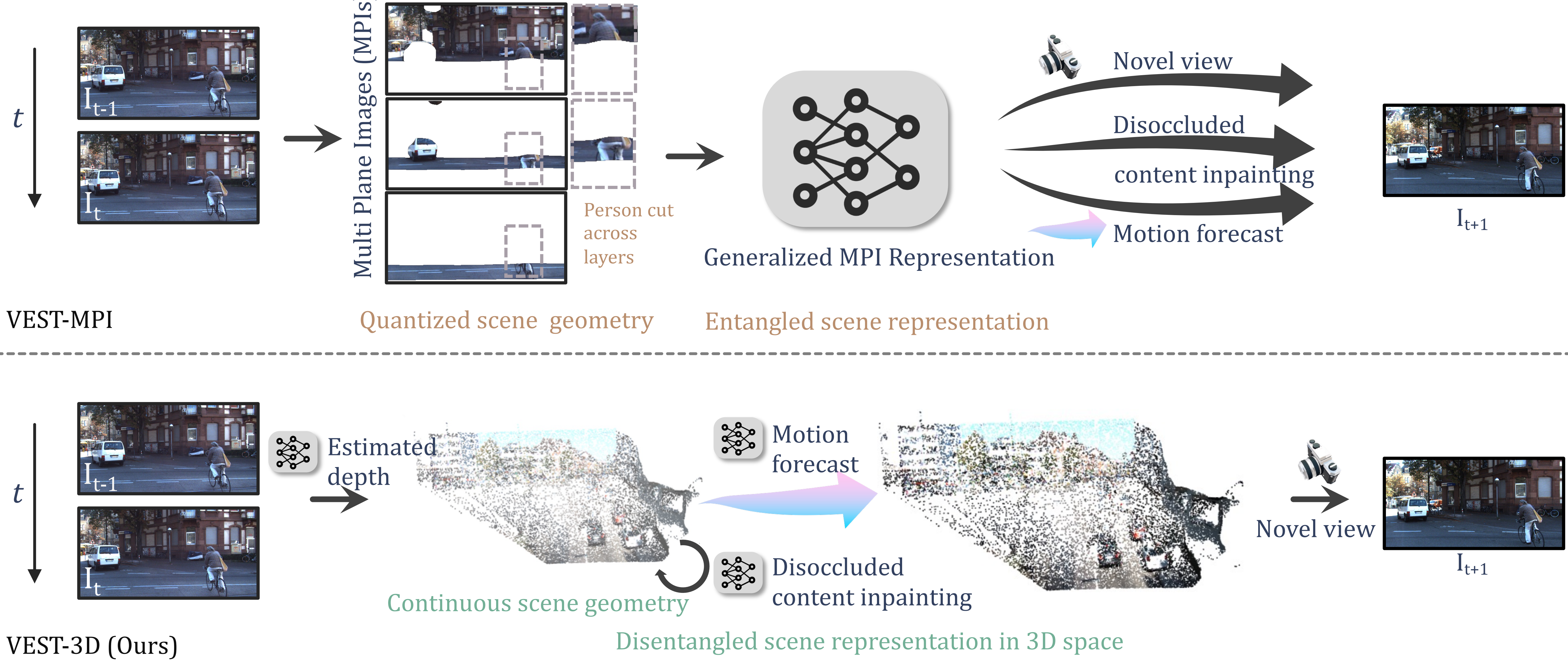
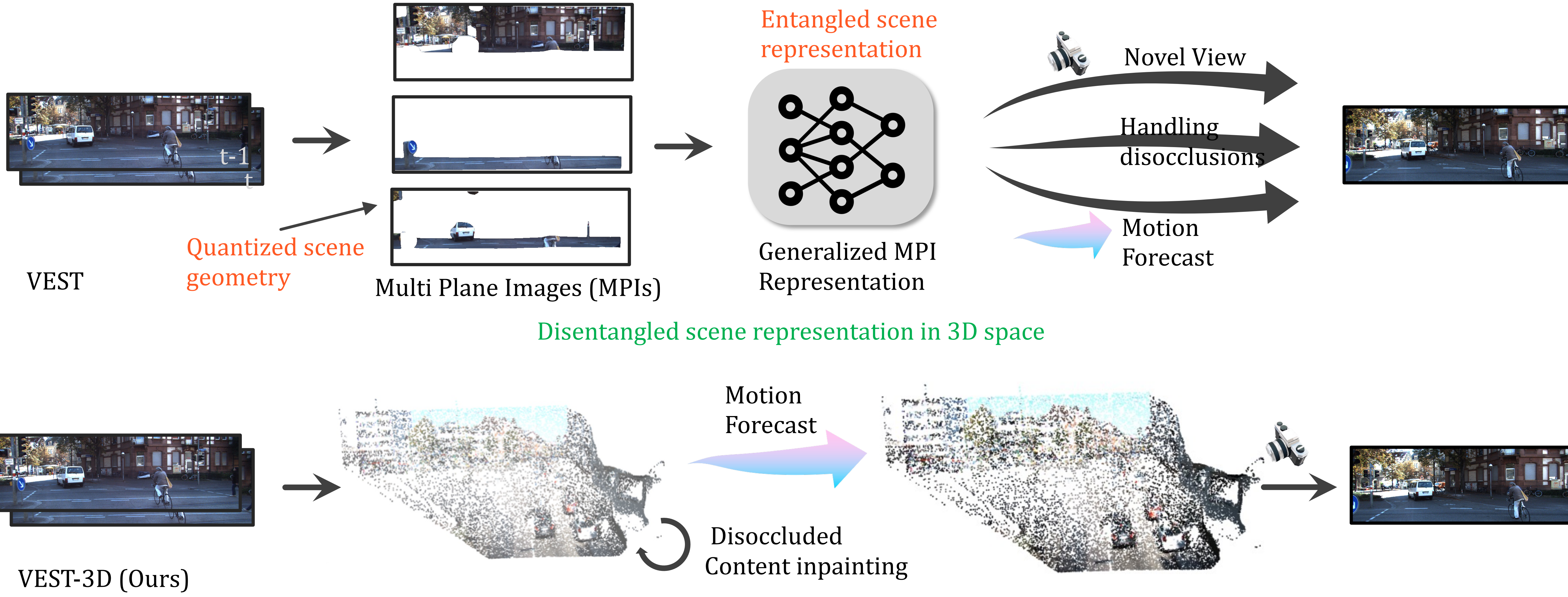
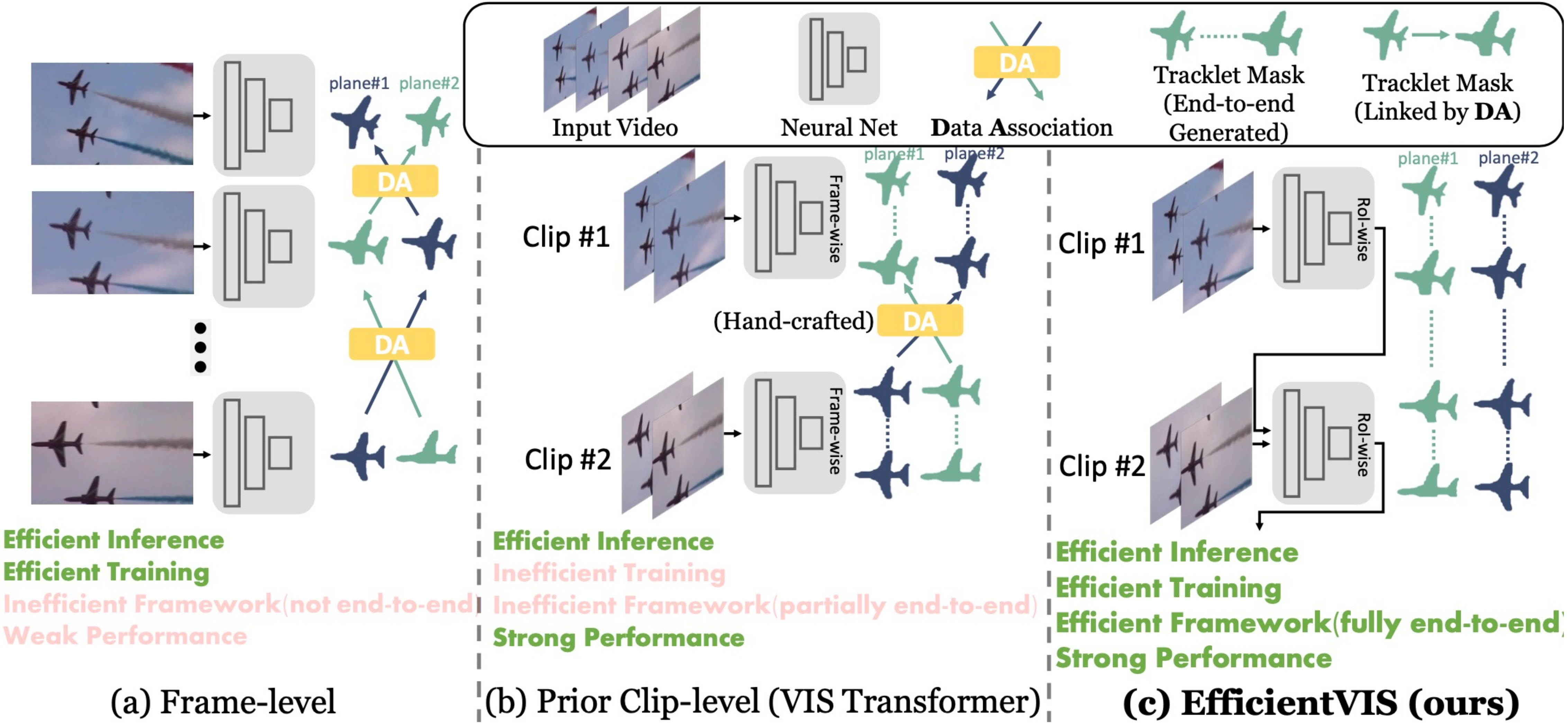

t
I_{t-1}
I_t
Generalized MPI Representation
Entangled scene representation
Scene geometry
Disoccluded content inpainting
Motion forecast
Entangled motion forecast
Multi Plane Images (MPIs)
Person cut across layers
Novel view
I_{t+1}
VEST-MPI
t
I_{t-1}
I_t
Estimated depth
Ego-motion forecast
Object motion forecast
Disentangled motion forecast
Disoccluded content inpainting
Disentangled scene representation in 3D space
Novel view
I_{t+1}
Ours


t
I_{t-1}
I_t
Generalized MPI Representation
Scene geometry
Disoccluded
content inpainting
Motion forecast
Multi Plane Images (MPIs)
Person cut across layers
Novel view
I_{t+1}
VEST-MPI
Entangled scene representation
Quantized scene geometry
t
I_{t-1}
I_t
Estimated depth
Motion forecast
Disoccluded content inpainting
Novel view
I_{t+1}
Continuous scene geometry
Disentangled scene representation in 3D space
Ours


t
I_{t-1}
I_t
Multi Plane Images (MPIs)
Person cut across layers
Generalized MPI Representation
Novel view
Disoccluded content inpainting
Motion forecast
I_{t+1}
VEST-MPI
Quantized scene geometry
Entangled scene representation
t
I_{t-1}
I_t
Estimated depth
Motion forecast
Disoccluded content inpainting
Continuous scene geometry
Novel view
I_{t+1}
VEST-3D (Ours)
Disentangled scene representation in 3D space


t-1
t
VEST
Quantized scene geometry
Multi Plane Images (MPIs)
Entangled scene representation
Generalized MPI Representation
Novel View
Handling disocclusions
Motion Forecast
Disentangled scene representation in 3D space
VEST-3D (Ours)
Motion Forecast
Disoccluded Content inpainting



(a) Frame-level

(b) Prior Clip-level (VIS Transformer)

(c) EfficientVIS (ours)